Submission 16

Compilation of Comments on Draft PATH Weight of Evidence Report

John Williams, Steve Smith, Tom Cooney, and Chris Toole

July 27, 1998

Thank you for the opportunity to comment on the draft PATH Weight of Evidence (WOE) Report, dated July 3, 1998. Unless otherwise stated, references to pages, figures, and tables refer to this report. This memorandum is a compilation of comments from the individuals listed above. Because only two NMFS biologists can attend the WOE meeting later this week, we request that, to the extent possible, discussions associated with comments 2.b and 4.a.2 occur on Thursday, when Chris Toole and Tom Cooney will be attending, and that discussions related to 4.a.1, 4.a.3, and 4.a.4 occur on Friday, when Steve Smith and Tom Cooney will be attending.

There are several attachments to these comments, which are included in separate files but which are intended to be directly appended as hard copies.

Attachment 1:	Spreadsheet ATTACH01.XLS for jump in survival table This supplements information presented in Comment 2.a.
Attachment 2:	WordPerfect document ATTACH02.WPD and Excel Spreadsheet ATTACH02.XLS. This supplements information in Comment 2.b.
Attachment 3:	Spreadsheet ATTACH03.XLS. This supplements information in Comment 2.b.
Attachment 4:	WordPerfect document ATTACH04.WPD. This supplements information in Comments 4.a.1 and 4.a.4.
Attachment 5:	WordPerfect document ATTACH05.WPD and spreadsheet ATTACH05.XLS. This supplements information in Comment 4.a.2.

1. Organization and Structure

In general, the document appears well-organized and structured in a manner that will facilitate evaluations of evidence. Two things that might improve the structure are: (1) a reduced discussion of the results and the influence of particular hypotheses on the results; and (2) a summary (maybe in outline form) of the specific hypotheses and the criteria and evidence for each one, with page numbers as a guide. The first comment is based on a recognition that those of us in PATH who saw earlier results needed to understand how they changed. However, there is no need for a weighting panel to be subjected to that detailed description of results and reading a detailed account of results will reduce the time available for them to focus on their task of evaluating information in support of alternative assumptions. The second suggestion is a way of making us concentrate on the evidence and criteria, rather than the results, and to lay out an easy guide for the weighting panel to follow.

2. <u>Results of Previous Analyses</u>

2.a. Jump in Survival at Beginning of Simulation Period: Need to Ensure That This Is Not a Systematic Error of Translating Retrospective Assumptions Into Prospective Runs

Criterion 4 presented in the draft report considers the validity of the method of projection. The report recognizes the difficulty in finding something to test against this concept, except for comparing projected short-term trends under different aggregate sets of assumptions against observed patterns in escapement in recent years (pages 57-58, 95-96). Data and projections for the Johnson Creek Index stock are analysed in the draft report, comparing a pair of aggregate hypotheses involving the alternative passage models to recent trends. The general conclusion reached is A*....spawning abundances projected by both aggregate hypotheses appear to be either unreasonably optimistic in their assumptions given the present condition of this particular stock, or both of them are excluding some source of mortality that has been operating in recent years*.@

Table 1 summarizes comparisons of recent trends in escapement against modeled short-term projections for a basic set of alternative aggregate hypotheses. The alternative aggregate hypotheses tested were generated by combining a set of five extra-mortality/climate hypotheses with two alternative sets of passage assumptions (FLUSH/FGE2/TURB5/PREM1 and CRISP/FGE2/?TURB4/PREM1). Details are included in **Attachment 1**. The median projected run in each case exceeds the recent historical median or average. The aggregate hypotheses involving FLUSH project out at an average of 52% higher than recent median escapements, the CRISP projections are 85% higher.

There is a consistent pattern in the relationship between projected median escapements and recent trends among aggregate hypotheses across stocks. The following table depicts that pattern in terms of the additional reduction necessary for the median projection to equal the median observed escapements.

Table 1. Percentage reduction required to equalize median projection with observed average escapement (1990-95)

Extra Mortality/ Climate Hypoth.	FLUSH/FGE2/ TURB5/PREM1	CriSP/FGE2 TURB4/PREM1
BKD/Markov Clm	31%	38%
BKD/Cyclical Clm	29%	37%
Regime Shift/Cyclical	31%	41%
Hydro/Markov	41%	55%
Hydro/Cyclical	38%	52%

Both modeling groups should look at the methods for translating retrospective assumptions into the prospective runs to ensure the pattern above is not the result of a systematic error. In addition, the possibitity that mortality rates may have increased in recent years should be directly addressed. For example, it has been suggested that predation by increased populations of avian predators in the lower river could cause significant mortality.

2.b. Purpose and Definitions of AKey Aggregate Hypotheses@: Aggregate Hypotheses Are of Limited Importance in This Exercise; To the Extent That They Are Important, the Proposed AHydro@ and ANon-Hydro@ Hypotheses Do Not Appear to Provide the Best Explanation of Alternative

Conclusions

The purpose of the weight of evidence task, as we understand it, is to provide evidence that can be used in assigning unequal weights to nodes in the decision tree for which all branches are currently weighted equally. Aggregate hypotheses are not required for this task. For purposes of efficiency, emphasis should be on those nodes exerting the greatest influence on alternative outcomes. The draft report identifies three non-drawdown nodes as those having the greatest influence on results: passage/transport model, distribution of extra mortality (alpha/delta), and linkage of extra mortality to other factors (extra mortality/climate hypotheses). Additionally, the draft identifies at least one drawdown node (the transition period) as being important for those outcomes involving A3. It is reasonable to concentrate on a comparison of alternative assumptions at each of these four nodes. If unequal weights are identified by the weighting panel for any of these nodes, the decision analysis can be re-run and the results should be less uncertain than the current results. Alternative aggregate hypotheses are <u>not</u> required for this analysis or for the weighting process.

The only reason that we can see for identifying or reviewing aggregate hypotheses is to determine if any of the 1920 combinations of assumptions (i.e., 1920 possible aggregate hypotheses) are not internally consistent. These inconsistent hypotheses can then be eliminated from the mix, which should further reduce uncertainty in the results. We aware of no combinations of assumptions that are internally inconsistent. For example, we see no reason why either FLUSH or CriSP can=t be coupled with alpha or delta or any of the extra mortality assumptions, let alone with any of the passage or drawdown assumptions. If someone is proposing that there are inconsistent combinations of assumptions within the 1920 combinations under consideration, the reasoning should be explained and reviewed by the PATH group. The weighting panel may also want to address this issue since, in a sense, it is also a weighting exercise (i.e., it may be a matter of opinion whether aggregate A is more logical than aggregate B).

If this suggestion is not accepted and for some reason the group wants to focus on a limited set of aggregate hypotheses, then the purpose of this approach should be better described. One purpose of identifying aggregate hypotheses may be to describe the combination of assumptions that best distinguish among: (1) the best management action, based on relative performance of one management action compared to another; and (2) adequacy of each management action, based on their performance relative to an approximation of the quantitative components of the jeopardy standard that NMFS applied in the 1995 FCRPS biological opinion (Ajeopardy standards@). **Note that these two questions need not be hierarchical: managers will want to know the performance of all management actions relative to the Ajeopardy standards,@ not just the performance of the best action.**

The draft report identifies two aggregates for consideration: AHydro@ (CRiSP+PROSPA+Regime Shift EM) and ANon-Hydro@ (FLUSH+PROSPD+Hydro EM). We question whether these particular aggregate hypotheses best explain the results. If a limited number of aggregate hypotheses are to be evaluated (which, to repeat, we don=t support), then
we propose two alternative methods for determining which aggregate hypotheses best distinguish among alternative management actions and the ability of those actions to meet the Ajeopardy standards.@ <u>In the first approach</u> (**Attachment 2**), we propose that there are two issues with repect to explanatory power of an aggregate hypothesis. One is: if this hypothesis is true, how likely is it will result in a particular outcome? An example would be, if you choose a complex set of assumptions
FGE1+TURB5+PREM1+PROSPA+Regime EM+PRER2+EJUV1+TJUVb <u>vs</u>
FGE2+TURB4+PREM3+PROSPD+Hydro EM+PRER1+EJUV2+TJUVa as the aggregate hypothesis and its complement to explain whether A1,A2 or A3 performs better relative to each other, there is a 100% likelihood that the first aggregate would result in A1,A2>A3 and that its complement will result in A3>A1,A2. The problem is that these aggregate hypotheses represent only two of 1920 possible aggregate hypotheses and many alternative hypotheses could also provide the same results. In fact, the first aggregate

explains only 0.9% (1/114) of the possible combinations that result in A1,A2>A3 and its complement explains only 0.06% (1/1806) of combinations resulting in A3>A1,A2. Therefore, a second criterion is necessary - the aggregate hypotheses under consideration should, in addition to having a high likelihood of distinguishing among alternative outcomes, explain a high proportion of the possible combinations that produce a given outcome. To meet the second criterion, the number of assumptions wrapped into the aggregate hypothesis must be reduced.

Choice of the Abest@ aggregate hypothesis involves optimizing the two criteria: likelihood that the aggregate will produce a given outcome and proportion of combinations resulting in the same outcome that are explained by the aggregate. **Attachment 2** describes the performance of various aggregate hypotheses relative to these criteria. For example, if one wanted to distinguish between the alternative outcomes of AA2 Passes Jeopardy Standards@ vs AA2 Fails,@ it appears that the ANon-Hydro@ and AHydro@ aggregate hypotheses are a poor means of distinguishing among these outcomes. The ›Non-Hydro@ aggregate explains only 6% (5/85) of AA2 Passes@ outcomes and, with this hypothesis, there is less than a 50-50 chance (42% [5/12]) that AA2 Passes@ will result. A better predictor is CRiSP+HYDRO EM, which has a 100% likelihood of resulting in AA2 Passes@ and explains 56% of the AA2 Passes@ results. The complementary hypothesis, FLUSH+(Any EM Other than Hydro EM), has a 100% chance of predicting AA2 Fails@ and explains 46% of the AA2 Fails@ results. CRiSP+HYDRO EM is also a better predictor (for both likelihood and proportion of combinations represented) of AA3 Passes@ than is the AHydro@ aggregate hypothesis. Also, one-factor hypotheses perform nearly as well as the two-factor hypotheses relative to likelihood and they perform much better relative to proportion of results explained.

We are not recommending particular aggregate hypotheses as the Abest@ using this method, since there may be reason to weight one or the other criterion more strongly. Rather, we suggest that, if PATH chooses to pursue identifying aggregate hypotheses to be emphasized in the weight of evidence approach, this approach for describing which hypotheses best explain the results should be included. In particular, PATH should review this information to determine if other aggregate hypotheses are more appropriate than the AHydro@ and ANon-Hydro@ hypotheses.

<u>The second method</u> that we suggest for describing aggregate hypotheses involves examining frequency distributions for individual stocks that are usually the Asixth best@ stocks. The modelling analysis conducted through PATH subjects each action to every possible combination of a set of specific elements. Comparisions across actions need to maintain consistency across the unique combinations of assumptions. The following questions are suggested as alternatives consistent with comparison across combinations of hypotheses:

A) What is the expected performance of each action against survival and recovery criteria?

B) Does one particular action consistently result in higher levels of returns across index stocks?

C) What is the relative difference in performance of the best action relative to other alternative actions?

Table 4-6 in the Evaluation section of the draft Weight of Evidence report lays out the projected escapement for the Johnson Creek index stock under scenario A1 against alternative combinations of assumptions. Tables 1 and 2 in **Attachment 3** depict the results of organizing information in a similar manner, consistent with the alternative questions, for two stocks (Minam and Sulphur Creek) that represent the majority of cases contributing to the sixth best stock category for the 48 year recovery criteria. After reviewing the tabulated results, the analysis was organized to depict factors that have a large differential influence. The results are organized into major blocks by passage model (CriSP and FLUSH) and prospective model (Alpha and Delta). Within each block, rows represent alternative climate/extra mortality hypotheses. The columns within each block represent two alternative actions; A2 and A3. Two variations are presented for Action A3. The Low variation includes pessimistic assumptions regarding drawdown implementation and equilibrium times, the High represents the optimistic assumptions. Each action is further split based on the assumed predation reduction (0 and 25%).

Several consistent patterns are apparent in the results, paralleling the conclusions reached regarding Johnson Creek projected escapements in the draft report. The first set covers both the survival and recovery criteria, the second and third set are specific to the 48 year recovery criteria and the 24 year survival criteria, respectively.

> Given the BKD hypothesis, the alternative climate assumptions made very little difference in the projections.
>
> The results for each combination were consistently higher given the Delta model relative the Alpha model.
>
> Action A3 consistently gave higher projections than Action A2 under all combinations of assumptions depicted.
>
> The predation reduction assumptions had a substantial impact on the projected performance under the Hydropower EM/Climate hypotheses, especially assuming the FLUSH/Delta model combinations.
>
> Delta model results are consistently higher for both actions, regardless of the choice of passage models or predator effectiveness assumptions.

<u>48 Year Recovery</u>

> The absolute differences between performance under A3 and that under A2 given specific combinations of assumptions were roughly 2-3 times higher under the FLUSH model than under CRiSP for the 48 Year recovery criteria.
>
> Action A3 exceeds the 48 year recovery criteria under all assumptions given the optimistic assumptions for drawdown implementation and equilibria. Under the pessimistic drawdown assumptions, the 48 year criteria are missed only under the BKD/Alpha/Low predator effectiveness combination for Minam (46% vs 50%).

<u>24 Year Survival</u>

> Results were generally more optimistic under the combinations including CRiSP than under those incorporating FLUSH.
>
> > Under CRiSP, Action A2 met the 24 yr criteria with the exception of Sulphur Creek under the BKD and Regime shift hypotheses using either ALPHA or DELTA assumptions.
> >
> > Under FLUSH, Minam did not exceed the criteria under the BKD or Regime shift hypotheses given A2. For Sulphur Creek, only the High predation reduction combined with the hydropower EM hypotheses exceeded the criteria.

The approach described above appears to support focusing on the same set of key factors as the analysis presented in the draft report, except that predation reduction is included.

3. Evaluation Criteria

Regarding Criterion 1 (p. 4-5), we suggest that the intent of the need for Aclarity@ should be elaborated upon. In some sections of the draft report, it appears that Aclarity@ of a hypothesis is equated with Asimplicity@ of a hypothesis (see our Comment 4.a.3). We are concerned that, to the extent that factors influencing salmon survival are complex, this will put more complex hypotheses at a disadvantage relative to more simplistic hypotheses.

Regarding the questions that the criteria are designed to address (e.g., p. 19), we believe that all actions need to be evaluated relative to the jeopardy standards, not just the best action. (See our comment 2.b).

4. Evidence Associated With Each Major Hypothesis

4.a. Passage/Transportation Model

4.a.1. Passage models do not predict recent in-river survivals accurately

PIT-tag data are available beyond those on which the models were calibrated. Proportions of yearling chinook salmon released from the Snake River trap detected at McNary Dam from 1989 to 1992 and recently developed estimates of survival between Lower Granite and Bonneville Dams provide Aout-of-sample@ checks of the degree to which the passage models predict in-river survival under recent conditions.

Our analyses of these PIT-tag data led to the following results and conclusions:

> (1) In every year from 1989 to 1992, the CriSP model predictions were closer to observed data than the FLUSH model or a spreadsheet model we developed using survival estimates from the 1970's.
>
> (2) CriSP predictions were considerably lower than observed data in 3 out of 4 years; the lower flow years.
>
> (3) The best predictions were made by our spreadsheet model using survival estimates from the 1990's.
>
> (4) In the 3 lowest flow years, FLUSH predictions were closer to the spreadsheet model using the 1970's survival estimates than they were to the observed data.
>
> (5) Models used to project survival in the future must reflect that relatively high survival is possible (and has occurred in recent years) under low flow conditions in the Snake River.
>
> (6) Based on empirical data, we estimated survival from Lower Granite Dam to Bonneville Dam for wild yearling chinook salmon to be 45.5% in 1997 and 57.8% in 1998.
>
> (7) CriSP predictions matched our estimates better in the higher flow and spill years of 1997 (CriSP: 43.7%) and 1998 (51.5%) than in the lower flow years 1989-92. We did not receive

FLUSH predictions for 1998.

4.a.2. Estimation of Historical Transport Experiment Control Survivals - Likelihood of FLUSH vs CRiSP Estimates CAN Be Inferred From Reach Survival Estimates for Certain Years

The draft WOE report concludes that the primary difference between CriSP/T4 and FLUSH/T1 is the estimation of survival of control fish in transport experiments, particularly in years prior to 1980. However, the draft report is very pessimistic about the ability to distinguish which of these assumptions is most likely. Page 62 concludes that reach survival studies are not very helpful in evaluating the difference between FLUSH s and CriSP Vcl transport control survival estimates. We disagree and suggest that it is possible to infer that certain estimates of transport control survival are very difficult to reconcile with reach survival estimates.

For example, the greatest discrepancy among FLUSH and CriSP estimates of transport control survival occurred in 1971 (Figure 3-23). Raymond's (1979) reach survival estimate for the two-project Little Goose to Ice Harbor reach in 1971 was 0.48 (Marmorek and Peters 1998, Table A.2.1-2), or approximately 0.69 per project. The FLUSH s estimate for transport control survival through the 6+ project reach from Little Goose to Bonneville (Marmorek and Peters 1998, Table A.3.1-1) is 0.65 (spreadsheet CONTSERV.XLS from C. Peters; see also Fig. 3-23), or approximately 0.94 per project. This FLUSH estimate of much higher survival through a considerably longer reach does not appear to be consistent with the Raymond (1979) estimate. Note that the FLUSH s estimate includes adjustment for mortality of control fish passing through Little Goose twice as a result of the experimental procedure (Table 3-9). On the other hand, the CriSP 6+ project transport control estimate for 1971 is 0.082, or approximately 0.68 per project, which is nearly identical to the Raymond mean per-project estimate.

Attachment 5 indicates that FLUSH average per-project transport control survival estimates are at least 10% higher than Raymond run-at-large estimates in most years for which comparisons are available (exceptions 1978-80). CriSP per-project estimates are within about 2% of Raymond's estimates in all years except 1976 and 1979, when they are lower than Raymond's estimates.

We admit that mean per-project survival estimates are a crude way of comparing survival through different reaches, but this method appears to be pointing out some fairly large discrepancies. Two points to consider:

(1) If mean per-project transport control and run-at-large survival estimates don't match up, we would expect the transport controls to experience lower survival because of the increased handling of the transport controls, compared to the run at large. Also, return rates of fish in transport experiments (whether controls or transported fish) are generally an order of magnitude lower than SARs of the run at large (**Attachment 5**).

(2) An obvious question arises: if both models fit the Raymond reach survival estimates well when diagnostics were performed (Fig. 4-3), then why is this discrepancy occurring? Estimates of Vn and Raymond reach survival can be used to estimate survival outside of the Raymond reach for each model (**Attachment 5**). In 1971, if we have used the available information correctly, FLUSH inriver migrants must have had 100%+ survival from Ice Harbor to Bonneville, for both the estimate submitted for the reach survival diagnostics and the Vn estimate to be true. It is possible that we are mixing and matching available information incorrectly, but we suggest that the conclusions that both models fit the Raymond reach survival estimates equally well be revisited and the results submitted for passage model reach diagnostics be compared in more detail against Vn estimates.

4.a.3 Comments related to Section 4.2.1.3 Transportation Models (Table 4-6 and 4-7).

Clarity

FLUSH/TRANS1

B*In spite of alignment with research results, it is not justifiable to use T/C values from historic truck transportation research as a basis for the prospective modeling as nearly all fish are now barged.   As an indication of problems with trucks, in 1978 trucked fish from Lower Granite Dam returned at 0.013% rate; whereas, those barged from Lower Granite Dam returned at 0.116% rate. The ISAB Report 98-2, dated 27 February 1998, evaluated the efficacy of transportation and concluded* **A***Trucks should not be used in the transportation program due to a lack of information needed to advise management, due to an absence of current research programs to collect such information, and because historical indications on truck transport are negative.***@**

--{{Counter argument to Paul=s comment that in 1996 NMFS wrote a memo to Mundy where NMFS agreed on a T/C data set that included trucked fish. That was prior to the Hydro group deliberations in 1997. After careful inspection of the data and new analyses based on the Hydro Group discussions, it became clear that problems occurred with the trucked studies related to release sites on the shoreline downstream of Bonneville Dam. NMFS (since mid-1997) recognizes a problem with trucked studies and does not now accept truck-based T/C values for use in models to derive T/C values from which to draw for prospective studies.}}

B*Fish condition has improved in recent years. Data clearly indicate that delayed mortality has decreased since the early to mid-1970s (Williams and Matthews 1995). Descaling has also decreased substantially at dams as improvements at the dams were made (see Figure !, submission 22.) SARs of Snake River fish jumped substantially in 1982 with the installation of the debris boom at Lower Granite Dam and remained higher than most years in Period II through 1985 (see figure 4-10a). Outside of 1971 (debris conditions in the forebay of the dam are unknown), the only year with a high SAR in Period II was the first year Lower Granite Dam came on line and no debris had accumulated in front of the powerhouse.*

CRiSP/TRANS4

B*There is no apparent biological reason to make a break between pre-1980 and post-1980 for historical and current periods. Independent of descaling values observed at dams collected during research efforts, the years with high SARs between 1969 and 1985 (a nearly, to fully completed hydropower system) all occurred under conditions where little to no debris existed at the uppermost powerhouse on the Snake River (1969, 1970, 1975, 1982-85). Further, as indicated above, it is inappropriate to use studies based on truck transportation.*

Table 4-7 Mechanism

FLUSH/TRANS1

B*Values of T/C were drawn from all studies, including those with trucks. At a minimum, only*

T/Cs from barge studies represent current and future transportation conditions. Further, as detailed in submission 16, it appears that the FLUSH model underestimates survivals in many years, particularly those with lower flows, under present hydropower conditions. This has the effect of decreasing D.

B*According to results shown in Table D-3 (top graph), system survival is directly related to survival of downstream migrant fish (through the flow/travel-time/survival relationship.) Independent of the percentage of fish transported, system survival decreases with decreasing survival of non-transported fish (directly related to travel time.) This model does not explain a mechanism of why transported fish mortality decreases as survival of non-transported fish decreases. Figure D-3 indicates that survival of transported fish would decrease as much as 50%.*

CRiSP/TRANS4

B*Although descaling likely had large impacts on fish during the many of the years in the 1970s (see Williams and Matthews 1995), it is not clear how descaling would have differentially impacted transported fish compared to downstream migrants. Further, the debris in front of dams appeared to have some added effect on smolts as SARs in almost all years when large amounts of debris were in the forebay returned at lower rates than in other years (particularly between 1969 and 1985 - see details in submission 16). Although descaling rates at Lower Granite Dam decreased after installation of the trash boom in 1982, high descaling rates were occasionally observed at Little Goose Dam until the system was rebuilt with larger passageways in 1991/92.*

Consistency with empirical evidence
FLUSH/TRANS1 AND CRiSP/TRANS4

B*As identified in clarity above, both models used results from truck transportation. This is inappropriate as a basis for prospective models. The years when transportation research studied the efficacy of barges were 1978, 1979, 1986, 1989, 1995, and 1996. Not all of these studies apparently carry the same weight based on reviews of transportation studies by Mundy et al. 1994) which suggested that since control fish were trucked to release sites in 1978, 1979, 1986, and 1989, that they may not represent* Atrue@ *controls.*

Transport:Control ratios

Rigor:

(i) Applicability --SCORE LOWER THAN 3 FOR TRUCK STUDIES
Transport studies conducted with trucks are not applicable to conditions that currently exist within the hydropower system and do not represent conditions that would occur prospectively

(ii) There is NOT agreement on which data to use. NMFS does not agree with use of the truck transport studies as measures of the value of transportation.

(iii) Citing personal communications with Kiefer, IDFG related to the possibility that transport TCR are close to 1:1 is a selective use of data. Table 1 below provides complete TCRs from 1993 to 1995 and those derived from the 2-ocean returns from the 1996 outmigration. There is a wide range in TCRs of transported fish compared to those never detected in the system. Further, if a belief exists that old transport studies have little value because of handling of controls that were transported (as detailed by Mundy et al. 1994), then only the 1995 transport study has any validity in the PATH process.

Spawner-Recruit Data

The basic hypothesis of low D values is NOT IN ALL CASES consistent with mu over the range of data. The FLUSH model predicts low Ds in all years that survival of downstream migrants was low with assumed levels of TCRs that are in the 2:1 range. The CRiSP model assigns low Ds to periods when descaling was considered high and higher Ds when descaling was considered low. In the CRiSP model, levels of descaling are assumed low between 1980 and 1992 and high prior to 1980. In the FLUSH model, survivals of downstream migrants are consistently low from 1972 to the late 1980s, while TCRs over the period were generally ranged near 2:1. However, mu was low in 1982 to 1985 and high from 1986 to 1992. Further, neither model predicted the low mu that occurred in 1975 when the TCRs were low, downstream migrant survivals low, and descaling high. Finally, although WTT time is the main component in the FLUSH model to predict downstream migrant survival, WTT was NOT a good predictor of mu (Deriso et al. 1996, page 5-17).

Smolt-to-Adult Return Rates

It is a specious argument to use low SARs of transported fish in recent years to suggest that transportation is not working.

From the juvenile outmigration that occurred in 1969 (6 dams in place), 1970 (7 dams in place) and 1975 (8 dams in place), SARs were approximately 4% (see Figure 4-10a). Juvenile survival through the hydropower system was estimated at approximately 28, 16, and 18%, respectively in the 3 years. No transportation occurred in those years. Juveniles passed downstream only through spill and turbines. Spill occurred when flows exceeded the powerhouse capacity of approximately 60 kcfs and was less than 50% until flows exceeded 120 kcfs. Through much of the migration in each year, more fish passed through the turbines than through spill.

In contrast, from 1993 to 1995, estimated SARs of PIT-tagged juveniles marked above Lower Granite Dam that were subsequently not detected in the hydropower system (therefore, at a minimum, did not pass through collection systems at Lower Granite, Little Goose, Lower Monumental, or McNary Dams) ranged from 0 to 0.41 % - a MAGNITUDE decrease in SARs from 1969, 1970, and 1975. Of these fish, approximately 50% or less passed through turbines, the rest through spill. The estimated juvenile survival of these fish ranged from approximately 31 to

48%.

Thus, even with a nearly doubled increase in survival of fish that passed downstream through the hydropower system in recent years compared to 1969, 1970, and 1975, there was nearly 10-fold decrease in SARs. This data overwhelmingly indicates that other factors, in addition to the hydropower system, are responsible for low SARs. It is NOT an issue of transportation.

Table 1. Smolt to adult return rates of wild spring/summer chinook salmon PIT-tagged and released above Lower Granite Dam, 1993-1996 .

Year	Estimated LGR & LGO juv. transport	Adult rtn	SAR	Estimated non-detected (juv.)	Adult rtn.	SAR	Ratio of trans. to non-detected TCR
1993	2,764	3	0.11	1,052	0	NA	NA - no non-det.
1994	2,041	12	0.59	2,858	6	0.21	2.81:1
1995	2,506	9	0.36	2,193	9	0.41	0.88:1
1996 (2-ocean only)	416	2	0.48	1,888	5	0.26	1.82:1

4.a.4. Alpha vs Delta Distribution of Extra Mortality

Common year effects

pages 69-70: Evidence for inclusion or exclusion of common year effects is well laid out and complete. Physiological state upon arrival in the estuary might be added to "genetic composition" as a factor that might distinguish Snake River stocks from lower river stocks despite similar arrival timing.

General model formulation

pages 68-69: There is a confusing argument against the alpha model based on the combination of anthropogenic and environmental influences on the alpha parameter. Ignoring specific implementations of the alpha model, are models that include interactions between anthropogenic and environmental factors to be discounted as a general principle? Isn't it likely that anthropogenic and environmental factors interact in reality? "Simplicity" is equated with "clarity" in other sections pertaining to Criterion 1 (e.g. line 47, page 53), reflecting an apparent reductionist tone that pervades much of the document. "In biology, Occam's razor cuts your throat."

Validity of Method of Projecting Hypotheses

The methods of projecting the hydrosystem extra mortality hypothesis under the alpha and delta models are introduced on page 75, but much of the evidence needed to assess the validity of the methods is on pages 90-92 and in Submission 3 (Hinrichsen and Paulsen) of the Appendix. We agree with Hinrichsen and Paulsen that there are serious problems with the way the models project the hydro hypothesis. We are most concerned about the unfalsifiable relationship implemented under the delta model. The implementation of this hypothesis is likely to be critical to the outcome of the decision analysis, and the method employed at present appears guaranteed to give high ratings to management actions that have large effects on direct survival in the migration corridor. The available data do not appear to support the assumption that extra mortality for non-transported smolts is proportional to their in-river mortality, and the arguments deployed on pages 90-92 to rescue the method from the questions raised in Submission 3 are not convincing. For example, how is the statement "it follows that (λ_n) should be inversely related to V_n " (line 23, page 90) reconciled with the assumption that $(1 - \lambda_{n,r})$ is proportional to $(1 - V_{n,r})$, stated just 8 lines earlier?

4.a.4. Extra Mortality Hypotheses: Incorrect Assumptions About What Conditions Presently Exert the Largest Impact on Mortality of the Stocks

Hydropower, "BKD", and climate are the three main hypotheses that have received the most attention to date in the PATH process as factors that may have the largest influence on stock performance. However, most of the analyses have used the hypotheses singly to explain observed stock performance during the period of BY 1975 to 1990 . This has simplified the modeling efforts, but may have masked changes in conditions over the larger span from pre-hydropower system completion (prior to 1968 or BY 1966) to the present. It is clear that the construction and operation of dams had a large effect on the downstream migration of Snake River stocks in many of the early years after their construction. However, it seems equally clear that the dams did not have much of an effect on SARs in other years.

There was little to no decrease in SARs (Figure 4-10a Weight of Evidence Report) the first year after John Day Dam and each of the Snake River dams were completed (1968/69 for John Day Dam, 1969 for Lower Monumental Dam, 1970 for Little Goose Dam, and 1975 for Lower Granite Dam). Further, in 1982 SARs rebounded from previous low levels, dropped in 1983 and 1984 and then rose again in 1985 to levels nearly equal to those that occurred when only four dams existed. Thus, between 1968 and 1985 (hydropower system nearly complete, or completely in place) 7 (39%) of the 18 years had SARs that were equal, or nearly equal to historic SARs[1] (analyses of Columbia and Snake River stocks with the Delta model also showed low μ:s during these 7 years) (Submission 9 in Appendix).

It appears likely that multiple factors have affected the Snake River stocks between 1964 and

[1] The 7 years of relatively high SARs all had one thing in common with historic conditions when only 4 dams were in place: little to no debris existed in the forebay of the first dam encountered by migrating juvenile fish. In 1969, 1970, and 1975 debris did not exist because it was the first year that the dams were operated and debris had not yet collected. Similarly, a debris boom was placed in the forebay of Lower Granite Dam in 1982 and debris was removed from the face of the powerhouse. Yet, SARs decreased in 1986 (BY1984) and have remained low since, despite the fact that debris was no longer an issue.

present and that single hypotheses (even when used in an aggregate form) do not explain the observed changes in stock performance over time. To adequately address extra mortality that has affected the stocks over the entire period will require consideration of multiple-factor hypotheses such as the one proposed in Submission 11 in the Appendix of the Weight of Evidence Report. In further developing such an analysis, attention should be given to additional factors that may have affected extra mortality in recent years (e.g., increases in populations of marine mammals and shad).

5. Miscellaneous Comments

5.a Multifactor Hypothesis (page 86 & App. Submission 11): Bird Predation Component

Piscivorous bird populations in the lower Columbia estuary have increased substantially in abundance during the juvenile migration period for anadromous salmonids (e.g., Roby & Currins, draft 1998). Populations of Caspian Terns and of Double Breasted Cormorants have both increased. The establishment of tern and cormorant colonies in the lower Columbia is relatively recent, information has been summarized in Roby & Currins (draft). Cormorant populations have rebounded beginning in approx. 1987, following severe declines around the turn of the century. The existence of a Lower Columbia. Caspian Tern colony was first cited in 1984 (1000 pairs on East Sand Island). By 1987, the colony had begun shifting to Rice Island (1,500). By 1991 the numbers had increased to 6200 breeding pairs, with a further increase to 8000 pairs by 1997. Cormorant colonies were discovered on East Sand in 87-88, and on Rice Island in 88. The most recent estimates of population levels are for 1997, approximately 6400 breeding pairs were surveyed on the two islands.

Feeding studies combined with bioenergetic calculations and PIT tag samples from the vicinity of the major Caspian Tern colony on Rice Island all support substantial impacts by these birds on migrating salmonids. Given the trends in population size described above, the potential impact of avian predators in the estuary increased substantially from 1987 through the 1990's.

The effects of increased avian predation are currently not directly included in the prospective modeling. Anderson (Attachment 11) has suggested a mathematical model designed to accomodate a retrospective assessment including avian predation. There are some problems with that approach including: a) avian predator populations increased in the later years of the available data series used for assessing mortality patterns - the lack of contrast will be a problem in developing meaningful coefficients from this particular assessment; b) the temporal pattern in the avian predation increase mimics the later stages of the trend in hatchery production suggesting the two factors could be confounded; c) the effects of bird predation in the lower estuary may not be primarily on transport releases - it is not clear how the Anderson model would allocate increased impacts in recent years to the various survival and productivity terms and whether that allocation would be biologically realistic.

Nevertheless, the evidence for increased bird predation in recent years is substantial. The analyses of the potential effects of the Snake River hydropower actions should take this factor into account. At a minimum, sensitivity analyses that would allow decision makers to gauge the potential response to the Snake River actions under two scenarios; bird predation returned to 1970's-early 80's levels and predation rates continuing at recent increased levels. Rough estimates of the potential rates can be calculated from existing information, additional surveys of PIT tags deposited by the birds on Rice Island could contribute to the estimate and may be available this fall.

5.b. Other Miscellaneous Comments

P. 10 Comment that CriSP and FLUSH Vn=s are similar for A3 ignores four years (of the 16 years, or 25%) for which ranges of estimates from each model do not overlap. Results may be similar in moderate and high flow years, but they are dissimilar in low flow years.

P. 12 It is imperative that the HAB B hypothesis be implemented correctly before final results are presented this fall. At this point, we do not know if this hypothesis influences outcomes or not. If it does influence outcomes significantly, we will regret not having subjected it to the weight of evidence process along with other influential assumptions.

P. 24 New tables, such as 3-3 through 3-5, should be presented which show the results of A3>(1.05*A1,A2) to determine what influences Asignificant@ differences in relative performance of the management actions.

P. 26-27 The multifactor tables for performance relative to jeopardy standards that Calvin made available to us should be presented in final results. (Note: in keeping with Comment 3, the detailed results may be in the form of an appendix or a separate report, rather than occupying a significant part of the WOE document).

P. 29 and similar log-scale figures. Suggest second axis innormal units for those of us who can=t do logs in our heads. For example, it seems important to know that the 1.0 vs 1.5 log difference between FLUSH and CRISP in 1978 and 1979 represents a 2.7- vs 4.5-fold increase in productivity, whereas the differences in the 1980's and 90's represent only about a 1.3- vs a.6-fold difference in productivity.

P. 52 Table 4-1 mixes assumptions and evidence. As stated in main comments, we do not support pursuit of these aggregate hypotheses. If we are over-ruled, this table needs significant improvement as well as clarification of whether its purpose is to lay out the components of the aggregate hypothesis or to summarize the evidence for and against each component.

The definition of Asystem survival@ under footnote 1 does not match the definition in other locations. Part 2) should describe a scalar of transported:in-river SARs, not SAR of transported fish and should describe how that scalar is applied to 1).

P. 55-56 and elsewhere: What were the specific aggregate hypotheses that were compared to the spawner-recruit data and how were these results summarized in Table 4-2? This test needs more description. For instance, what was assumed for the PREM and FGE assumptions and how were these aggregated ? Are the EM assumptions irrelevant in retrospective analyses? How were the number of parameters determined and why are they the same for FLUSH and CRISP?

P. 60 Re: Criterion 1. Given the significant changes in both CriSP and FLUSH in recent years, why id the Barnthouse (1993) review relevant? The specific documentation used for each model should be described here. Are we just using the model component descriptions in other sections in other sections of the PATH report, or is there useful recent documentation of either model that a panel member can review today? We would not be able to direct anyone to a current description of FLUSH nor could we direct them to a description of the specific parameters fit to the general relationships described in the CRISP documentation on the internet.

P. 68 Top of page comment about internal consistency of aggregate hypotheses - in previous discussion, which of the aggregates are you referring to as less internally consistent and what is the definition of

Ainternally consistent?@

P. 79 The vertical lines in this figure are not in the right places - probably a result of re-sizing the figure. This also occurred in Fig. 1 of Submittal 1 in the appendix.

6. Response to Section IV Submission 22 "Passage Models and Reach Survival Data"

Steven G. Smith and John G. Williams

The comments in Section IV of Submission 22 (Wilson et al) regarding our analysis of observed 1989-1992 PIT-tag data (Submission 16) are largely irrelevant to the central point we were making. In fact, the comments serve to obscure the important points, which are restated here for clarity:

(1) PIT-tag data are available for 1989-1992. The *observations* we dealt with were the percentages of yearling spring/summer chinook salmon released from the Snake River trap that were subsequently detected at the Snake River dams. The percentages detected at McNary Dam, and the degree to which CRiSP and FLUSH predictions agreed with the observed percentages, were of particular interest.

(2) CRiSP predictions were lower than observed data in 3 of the 4 years, and FLUSH predictions underestimated observations in all 4 years. The largest discrepancy for both models was for 1992, which was the lowest flow year.

(3) FLUSH predictions were similar to those from a spreadsheet model we developed based on survival estimates obtained in low flow years in the 1970's. Observed data were more similar to results obtained when the spreadsheet model used (much higher) survival estimates obtained under low flows in 1994 and 1995.

(4) Several model parameters other than survival might contribute to the discrepancy, particularly FGE at Lower Granite, Little Goose, and McNary Dams. In fact, data on detections at Lower Granite and Little Goose Dams suggest that FGE values assumed in PATH are in error. However, no reasonable combination of assumptions for FGE at the dams can explain the entire difference between observed and predicted percentages detected at McNary Dam. **The passage models must be underestimating survival.**

The following are responses to specific statements made in Submission 22 regarding the analysis of 1989-1992 data.

The following quantities were described as "observed" in our submission: release dates, number released, numbers and percentages detected at Lower Granite, Little Goose, and McNary Dams, number of detected fish returned to the river at Lower Granite and Little Goose Dams, travel times of PIT-tagged fish in 1994 and 1995, FGE estimates from field studies, and flows in 1994 and 1995. **Survival estimates were not described as "observed."** Table 4 of our submission includes per-project survival probabilities which are identified in the table caption as those probabilities implied by (i.e., consistent with) observed data, predictions and assumed FGEs. One row of the table is labeled "Observed," but these are *not* "observed" survival estimates, but the survival probabilities that are consistent with the observed data and assumed FGEs.

The purpose of the exercise was not to develop reach-by-reach survival estimates for 1989-1992, as implied by the comments. Rather, it was to examine observed PIT-tag data, compare them to model predictions, and to discuss potential faults in assumptions (including incorrect FGE assumptions) that might explain discrepancies between model and observation. Our spreadsheet model made use of survival estimates from other years, but only to explore the range of predictions for the percentage of fish detected at McNary Dam. The spreadsheet model did not estimate survival.

We did not "report" (or even calculate) survival estimates for Snake River trap to Lower Granite Dam of 90% in 1991 and 77% in 1992. We said "If 46% FGE is correct...then survival of 90% and 77% is implied." We did "report" survival estimates exceeding 100% for 1989 and 1990 using this "method" because we explicitly stated that the assumed value of 35% FGE at Lower Granite Dam was erroneous, since the proportion of fish detected exceeded 35%. In any case, discussion of reach by reach "survival estimates" is irrelevant, as those were not the quantities we asked the modeling groups to predict.

Our description of the data from 1994 and 1995 that we used to inform our choice of survival probabilities for the spreadsheet model was inadequate. The value of 0.892 listed in Table 4 was derived from Lower Granite-to-McNary estimates (CJS model) for daily groups of yearling chinook salmon leaving Lower Granite Dam in 1995 (Appendix Table D3 in Smith et al 1998). Groups leaving between 13 and 25 April, inclusive, had flow exposure indices less than 90 kcfs and survival estimates based on 5 or more detections below McNary Dam. Their weighted (inverse variance) average LGR-MCN survival estimate was 0.633, which works out to 0.892 per project for the 4 projects. Passage indices were not used to weight the average, but this adjustment would make little difference. The exact value of this estimate was not important. That it was considerably higher than the estimate from the 1970's and appeared much more consistent with observed data was.

The "direct estimates of survival from release to McNary Dam" calculated by Wilson et al. differed considerably from those we were said to have "reported" (our "estimates" were higher). Inspecting Table 1 of Wilson et al. and Table 1 of our submission, the differences in "estimates" are almost entirely due to differences in the numbers of PIT-tagged fish counted as detected and returned to the river at Little Goose Dam in 1990, 1991, and 1992. Closer inspection suggests that Wilson et al. made an error in transcription from our table: their numbers of fish detected and returned at Little Goose Dam in 1990, 1991, and 1992 are identical to the line just below, the number of fish detected at McNary. If the correct numbers returned at Little Goose Dam (0, 0, and 1 in 1990, 1991, and 1992, respectively) are substituted into their table, their estimation method gives estimates of "Direct (FGE) Survival Rel to McNary" of 54.6%, 41.4%, and 82.0%, which work out to per-project estimates of 88.6% in 1990, 83.8% in 1991, and 96.1% in 1992. These estimates are nearly identical to the probabilities we calculated as being consistent with the observed data and the assumed FGE values.

FLUSH predictions are for seasonal averages; it is true that they cannot be fairly expected to match data from short periods of time within a migration season as well as they match seasonal averages. The FLUSH prediction of a little less than 2% McNary detection for fish released from the Snake River trap in 1992 should be taken, then, as the proportion of PIT-tagged yearling chinook salmon that would have been detected at McNary Dam in 1992 had PIT tagging at the trap been proportional to the population passing. Given that Snake River trap catch in 1992 was not representative of the population passing the trap, the FLUSH model might be making very accurate predictions for the population as a whole, yet underestimating the observed data. However, the period during which the trap was operating with high efficiency likely included the peak of the migration. It is difficult to see how fish marked during a 2-week period beginning in late April could have a 10.5% detection proportion at McNary Dam, while outside that period the detection proportion was so low that the overall percentage was less than 2%.

Finally, yearling chinook salmon migrating past the Snake River trap after May 10 represent the end of the migration. In our PIT-tag studies from 1993 through 1998 we have observed a suggestion of decreased survival for groups of fish marked and released at the end of the season. Thus, it is not surprising that only 2 of 52 (3.8%) of fish released after May 10 were detected at McNary Dam. We note, however, that this is still about twice the proportion that the FLUSH model predicted for the entire population, including those that migrated earlier under presumably more favorable conditions.

Below are comments and suggestions for the section of the Weight of Evidence report that deals with PIT-tag data from 1989-1992 and 1997, 1998.

Page 36 line 5: suggest replacing "detection probabilities" with "proportion of PIT-tagged fish detected" in section heading. "Detection probability" has a specific meaning in the CJS model, separate from the quantity we're dealing with here.

Page 36 line 13: begins "Neither of these data sets," while the section is just about 1989-1992 data.

Applicability section for the 1989-1992 data should note that the data are from a mixture of hatchery and wild fish.

Page 36 line 15: while observed data could be compared to model predictions of the proportions detected at each dam, we received from the modeling groups and analyzed only proportions detected at McNary Dam.

Page 36 lines 17-19: we made no estimates of survival and did not compare model outputs of survival estimates to "observed" estimates.

Page 36 lines 23-25: Comments appear to be addressing survival estimates that we did not make. There are no "NMFS estimates." The data are quite clear: release fish above, count how many show up below, compare to what the models predict.

Page 37 line 18: a revision of our original Submission 16 has been submitted to consider FLUSH outputs for 1997.

Page 37 lines 20-23: applicability should be 1 or 2. The data are for wild yearling chinook salmon from the Snake River drainage, migrating through nearly the entire hydrosystem. Applicability of 2 might be considered only because the data are not stock specific.

Page 37 lines 25-34: questions of methodology are properly handled under Clarity and Rigor, not Applicability.